DOCUMENT RESUME

ED 458 288 TM 033 452

AUTHOR	La Marca, Paul M.
TITLE	Alignment of Standards and Assessments as an Accountability Criterion. ERIC Digest.
INSTITUTION	ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-01-07
PUB DATE	2001-11-00
NOTE	4p.; Based on material originally appearing in "Practical Assessment Research and Evaluation."
CONTRACT	ED-99-CO-0032
AVAILABLE FROM	ERIC Clearinghouse on Assessment and Evaluation, 1129 Shriver Laboratory, University of Maryland, College Park, MD 20742. Tel: 800-464-3742 (Toll Free); Web site: http://ericae.net.
PUB TYPE	ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Standards; *Accountability; *Educational Assessment; Elementary Secondary Education; Scores; *State Standards; Test Construction; Test Interpretation
IDENTIFIERS	*Curriculum Alignment; ERIC Digests

ABSTRACT

This digest provides an overview of the concept of alignment and the role it plays in assessment and accountability systems. It also discusses methodological issues affecting the study of alignment and explores the relationship between alignment and test score interpretation. Alignment refers to the degree of match between test content and subject area content identified through state academic standards. It is highly unlikely that a single test can achieve a desirable degree of match, and this fact provides the rationale for using multiple accountability measures and points to the need for studying the degree of match at the test level and the system level. The two overarching dimensions of alignment are content match and depth match. Alignment can best be achieved through sound standards and assessment development activities. The analysis of alignment requires a systematic review of standards and a systematic review of test items/tasks. Alignment is not only a methodological requirement, but also an ethical requirement, as it would be a disservice to students and schools to judge academic achievement based on a poorly aligned system of assessment. The systematic study of alignment is time consuming and can be costly, but the benefits outweigh the costs. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 458 288

Alignment of Standards and Assessments as an Accountability Criterion.

ERIC/AE DIGEST

EDO-TM-01-07
November 2001

Paul M. La Marca
Nevada Department of Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
❑ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 033 452

ERIC Clearinghouse on Assessment and Evaluation

BEST COPY AVAILABLE

Alignment of Standards And Assessments as an Accountability Criterion

Paul M. La Marca
Nevada Department of Education

To make defensible accountability decisions based in part on student and school-level academic achievement, states must employ assessments that are aligned to their academic standards. Federal legislation and Title I regulations recognize the importance of alignment, which constitutes just one of several criteria for sound assessment and accountability systems. However, this seemingly simplistic requirement grows increasingly complex as its role in the test validation process is examined.

This paper provides an overview of the concept of alignment and the role it plays in assessment and accountability systems. Some discussion of methodological issues affecting the study of alignment is offered. The relationship between alignment and test score interpretation is also explored.

The Concept of Alignment

Alignment refers to the degree of match between test content and the subject area content identified through state academic standards. Given the breadth and depth of typical state standards, it is highly unlikely that a single test can achieve a desirable degree of match. This fact provides part of the rationale for using multiple accountability measures and also points to the need to study the degree of match or alignment both at the test level and at the system level. Although some degree of match should be provided by each individual test, complementary multiple measures can provide the necessary degree of coverage for systems alignment. This is the greater accountability issue.

Based on a review of literature (La Marca, Redfield, & Winter 2000), several dimensions of alignment have been identified. The two overarching dimensions are content match and depth match. Content match can be further refined into an analysis of broad content coverage, range of coverage, and balance of coverage. Both content and depth match are predicated on item-level comparisons to standards.

Broad content match, labeled categorical congruence by Webb (1997), refers to alignment at the broad standard level. For example, a general writing standard may indicate that "students write a variety of texts that inform, persuade, describe, evaluate, or tell a story and are appropriate to purpose and audience " (Nevada Department of Education, 2001 p. 14). Obviously this standard covers a lot of ground and many specific indicators of progress or objectives contribute to attainment of this broadly defined skill. However, item/task match at the broad standard level can drive the determination of categorical congruence with little consideration to the specific objectives being measured.

As suggested above, the breadth of most content standards is further refined by the specification of indicators or objectives. Range of coverage refers to how well items match the more detailed objectives. For example, the Nevada writing standard noted above includes a variety of specific indicators: information, narration, literary analysis, summary, and persuasion. Range of coverage would require measurement to be spread across the indicators. Similarly, the balance of coverage at the objective level should be judged based on a match between emphasis in test content and emphasis prescribed in standards documents.

Depth alignment refers to the match between the cognitive complexity of the knowledge/skill prescribed by the standards and the cognitive complexity required by the assessment item/task (Webb 1997, 1999). Building on the writing example, although indirect measures of writing, such as editing tasks, may provide some subject-area content coverage, the writing standard appears to prescribe a level of cognitive complexity that requires a direct assessment of writing to provide adequate depth alignment.

Alignment can best be achieved through sound standards and assessment development activities. As standards are developed, the issue of how achievement will be measured should be a constant consideration. Certainly the development of assessments designed to measure expectations should be driven by academic standards through development of test blueprints and item specifications. Items/tasks can then be designed to measure specific objectives. After assessments are developed, a post hoc review of alignment should be conducted. This step is important where standards-based custom assessments are used and absolutely essential when states choose to use assessment products not specifically designed to measure their state standards. Whenever assessments are modified or passing scores are changed, another alignment review should be undertaken.

Methodological Consideration

An objective analysis of alignment as tests are adopted, built, or revised ought to be conducted on an ongoing basis. As will be argued later, this is a critical step in establishing evidence of the validity of test score or performance interpretation.

Although a variety of methodologies are available (Webb, 1999; Schmidt, 1999), the analysis of alignment requires a two-step process:

1. a systematic review of standards and
2. a systematic review of test items/tasks.

This two-step process is critical when considering the judgment of depth alignment.

Individuals with expertise in both subject area content and assessment should conduct the review of standards and assessments. Reviewers should provide an independent or unbiased analysis; therefore, they should probably not have been heavily involved in the development of either the standards or the assessment items.

The review of standards and assessment items/tasks can occur using an iterative process, but Webb (1997, 1999) suggests that the review of standards precede any item/task review. An analysis of the degree of cognitive complexity prescribed by the standards is a critical step in this process. The subsequent review of test items/tasks will involve two decision points

1. a determination of what objective, if any, an item measures and
2. the items degree of cognitive complexity.

The subjective nature of this type of review requires a strong training component. For example, the concept of depth or cognitive complexity will likely vary from one reviewer to the next. In order to code consistently, reviewers will need to develop a shared definition of cognitive complexity. To assist in this process, Webb (1999) has built a rubric that defines the range of cognitive complexity, from simple recall to extended thinking. Making rubric training the first step in the formal evaluation process can help to reinforce the shared definition and ground the subsequent review of test items/tasks.

Systematic review of standards and items can yield judgments related to broad standard coverage, range of coverage, balance of coverage, and depth coverage. The specific decision rules employed for each alignment dimension are not hard and fast. Webb (1999) does provide a set of decision rules for judging alignment and further suggests that determination of alignment should be supported by evidence of score reliability.

Thus far the discussion has focused on the evaluation of alignment for a single test instrument. If the purpose of the exercise is ultimately to demonstrate systems alignment, the process can be repeated for each assessment instrument sequentially, or all assessment items/tasks can be reviewed simultaneously. The choice may be somewhat arbitrary. However, there are advantages to judging alignment at both the instrument level and the system level. If, for example, decisions or interpretations are made based on a single test score, knowing the test's degree of alignment is critical. Moreover, as is typical of school accountability models, if multiple measures are combined prior to the decision-making or interpretive process, knowledge of overall systems alignment will be critical.

Why is alignment a key issue

In the current age of educational reform in which large-scale testing plays a prominent role, high-stakes decisions predicated on test performance are becoming increasingly common. As the decisions associated with test performance carry significant consequences (e.g., rewards and sanctions), the degree of confidence in, and the defensibility of, test score interpretations must be commensurably great. Stated differently, as large-scale assessment becomes more visible to the public, the roles of reliability and validity come to the fore.

Messick (1989) has convincingly argued that validity is not a quality of a test but concerns the inferences drawn from test scores or performance. This break from traditional conceptions of validity changes the focus from establishing different sorts of validity (e.g., content validity vs. construct validity) to establishing several lines of validity evidence, all contributing to the validation of test score inferences.

Alignment as discussed here is related to traditional conceptions of content validity. Messick (1989) states that "Content validity is based on professional judgments about the relevance of the test content to the content of a particular behavioral domain of interest and about the representativeness with which item or task content covers that domain" (p. 17). Arguably, the establishment of evidence of test relevance and representativeness of the target domain is a critical first step in validating test score interpretations. For example, if a test is designed to measure math achievement and a test score is judged relative to a set proficiency standard (i.e., a cut score), the interpretation of math proficiency will be heavily dependent on a match between test content and content area expectations.

Moreover, the establishment of evidence of content representativeness or alignment is intricately tied to evidence of construct validity. Although constructs are typically considered latent causal variables, their validation is often captured in measures of internal and external structure (Messick, 1989). Arguably the interpretation of measures of internal consistency and/or factor structures, as well as associations with external criterion, will be informed by an analysis of range of content and balance of content coverage.

Therefore, alignment is a key issue in as much as it provides one avenue for establishing evidence for score interpretation. Validity is not a static quality, it is "an evolving property and validation is a continuing process" (Messick, p. 13). As argued earlier, evaluating alignment, like analyzing internal consistency, should occur regularly, taking its place in the cyclical process of assessment development and revision.

Discussion

Alignment should play a prominent role in effective accountability systems. It is not only a methodological requirement but also an ethical requirement. It would be a disservice to students and schools to judge achievement of academic expectations based on a poorly aligned system of assessment. Although it is easy to agree that we would not interpret a student's level of proficiency in social studies based on a math test score, interpreting math proficiency based on a math test score requires establishing through objective methods that the math test score is based on performance relative to skills that adequately represent our expectations for mathematical achievement.There are several factors in addition to the subjective nature of expert judgments that can affect the objective evaluation of alignment. For example, test items/tasks often provide measurement of multiple content standards/objectives, and this may introduce error into expert judgments. Moreover, state standards differ markedly from one another in terms of specificity of academic expectations. Standards that reflect only general expectations tend to include limited information for defining the breadth of content and determining cognitive demand. Not only does this limit the ability to develop clearly aligned assessments, it is a barrier to the alignment review process. Standards that contain excessive detail also impede the development of assessments, making an acceptable degree of alignment difficult to achieve. In this case, prioritization or clear articulation of content emphasis will ease the burden of developing aligned assessments and accurately measuring the degree of alignment.

The systematic study of alignment on an ongoing basis is time-consuming and can be costly. Ultimately, however, the validity of test score interpretations depends in part on this sort of evidence. The benefits of confidence, fairness, and defensibility to students and schools outweigh the costs. The study of alignment is also empowering in as much as it provides critical information to be used in revising or refining assessments and academic standards.

References

La Marca, P. M., Redfield, D., & Winter, P.C. (2000). State Standards and State Assessment Systems: A Guide to Alignment. Washington, DC: Council of Chief State School Officers.

Messick, S. (1989). Validity. In R. L. Linn (Editor), Educational Measurement (3rd Edition). New York: American Council on Education – Macmillan Publishing Comany.

Nevada Department of Education (2001). Nevada English Language Arts: Content Standards for Kindergarten and Grades 1, 2, 3, 4, 5, 6, 7, 8 and 12.

Schmidt, W. (1999). Presentation in R. Blank (Moderator), The Alignment of Standards and Assessments. Annual National Conference on Large-Scale Assessment, Snowbird, UT.

Webb, N. L. (1997). Research Monograph No. 6: Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1999). Alignment of Science and Mathematics Standards and Assessments in Four States. Washington, DC: Council of Chief State School Officers.

The author would like to acknowledge Phoebe Winter, Council of Chief State School Officers, and Doris Redfield, Appalachia Educational Laboratory, for their assistance in critiquing this manuscript. I would like to acknowledge the CCSSO SCASS-CAS alignment work group for preliminary work in this area.

This Digest is based on material originally appearing in *Practical Assessment Research and Evaluation*.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract ED99CO0032. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education.

Permission is granted to copy and distribute this ERIC/AE Digest.